International Journal of Scientific Engineering and Technology
Volume No. 4, Issue No. 8, pp: 443 – 446.
ISSN: 2277 – 1581
01 August 2015

Speaker Dependent Voice Recognition Using Discrete Wavelet Transform

Angelo A. Beltran Jr.[1], Ericson D. Dimaunahan[2], Donde A. Deveras
[1]Department of Electronics Engineering, Adamson University, Philippines
[2]Physics Department, Mapua Institute of Technology, Philippines
abeltranjr@hotmail.com, eddimaunahan@mapua.edu.ph, dondeveras@yahoo.com

***Abstract*—This paper presents the effective and robust method for the feature extraction of the speaker dependent voice recognition. The authors developed a simple Matlab program for this purpose where the article discrete wavelet transform theory had been used. The voice of set of speakers had been inputted on the database and the discrete wavelet transform calculates the properties and variables needed in order to verify correctly the speaker. Experimental results show that our method is very effective and the results are satisfactory and finally, the wavelet-based voice recognition system and its performance are discussed and highlighted.**

***Keywords*—Speaker dependent, Voice recognition, Discrete wavelet transform**

I. Introduction

Speech recognition is one of the rapidly fast pace digital signal processing work for which in our perspectives it has many real world engineering applications. It can be used to perform and improve tasks that are automatically set such for example the voice commands for security purposes like window opening, feeding a pet, shutting off the lights, etc. all of which can then be possibly replaced the manual and classical interactions of human into something. In the recent years, research and technological advances in artificial intelligence able to increase the rate of recognition for speech recognition such for example the methods of artificial neural networks, hidden markov chains or models, the use of fourier analysis, gabor transform, and many more. Nowadays, there are some available software packages for voice recognition where speech and also the simple discussion can be converted into text through the normal conversation although that aids the person with functional disabilities. Although, this breakthrough was succeeded, there are still much more needed to be done when it comes to the voice or speech recognition and hence, the authors consider this also as a good project and a challenge as well that the authors believe it can be a fruitful and rewarding project. The speaker recognition through the natural conversation seems to be a good work that is a very challenging because there are lot of variable that needs to be consider such for example the pitch, angles, frequency, amplitude, etc. and this variables do vary from different humans as they speak and it makes the problem complex when we look deep inside the certain specific sets of variables that needs to be evaluated like consonants, vowels, etc. For now, specifically, the objectives of this paper are the following: (a) to design the speaker dependent word or voice recognition system where the system will take an input word signal from a user, (b) to compare the signal with every entry in an already stored code book database to recognize the voice said after the noise/silence removal, and (c) to use discrete wavelet transform (DWT) algorithm for the feature extraction for the recognition of a particular type of word spoken in the system. This paper is specifically limited to a speaker dependent voice recognition system and the method therein presented. The paper is organized as it follows: The section II briefly presents the fundamentals of the speech recognition, and discrete wavelet transform, Under the Section III experimental results are carried out in order to verify also the effectiveness of the proposed method. Conclusion ends the paper at Section IV.

II. Speech Recognition and Wavelets

A. Speech Recognition

The speech recognition is a task of recognition of patterns where words, sentences, etc. are being analyzed and examined.

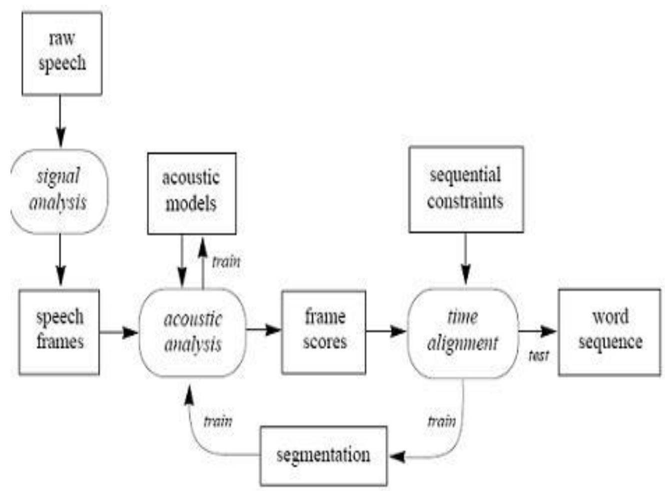

Fig. 1. A standard speech recognition system [8].

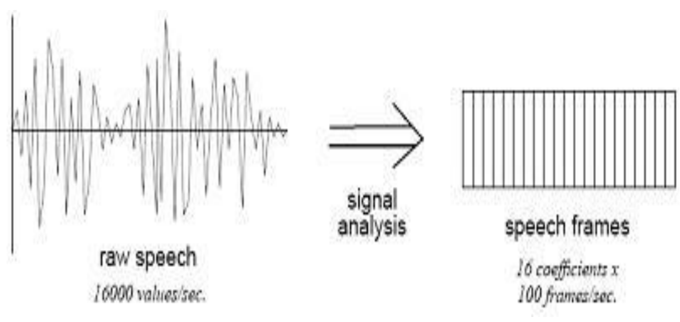

Fig. 2. The raw speech to frames [8].

 ISSN: 2277 – 1581
01 August 2015

B. Speaker Dependent and Speaker Independent [9]

Speaker dependent every persons voice are inputted and system is being trained which therefore achieves good accuracy for voice recognitions. In speaker independent, the prototype is now being train to identify and respond from voice of anyone.

C. Discrete Wavelet Transform

The daubechies wavelet is one of the popular wavelets and has been used for speech recognition [1]. Below are then the daubechies wavelets properties:

(a.) The support length of the wavelet function Ψ and the scaling function Φ is $2N-1$ The number of vanishing moments of Ψ is N (b.) Most dbN are not symmetrical (c.) The regularity increases with the order. When the N becomes very large, Ψ and Φ belong to $C\mu N$

A function $f(t) \in L^2(R)$ (defines space of square integral functions) can be represented as:

$$f(t) = \sum_{j=1-L} \{ \sum_{k=\alpha}^{-\alpha} d(j,k)\psi(2^{-j}t-k) + \sum_{k=\alpha}^{-\alpha} a(L,k)\varphi(2^{-L}t-k) \} \quad (1)$$

The function $\psi(t)$ is known as the mother wavelet while $\varphi(t)$ is known as the scaling function. The set of functions

$$\{ \sqrt{2}^{-L}\varphi(2^{-L}t-k), \sqrt{2}^{-j}\psi(2^{-j}t-k) \ \square\ j <= L, j,k,L \in Z \} \quad (2)$$

where Z the set of integers, is an orthonormal basis for $L^2(R)$ The numbers $a(L,k)$ are known as the approximation coefficients at scale L, while $d(j,k)$ are known as the detail coefficients at scale j The approximation and detail coefficients can be expressed as follows

$$a(L,k) = \frac{1}{\sqrt{2}} \int_{a}^{-a} f(t)\varphi(2^{-L}t-k)dt \quad (3)$$

$$d(j,k) = \frac{1}{\sqrt{2}} \int_{a}^{-a} f(t)\psi(2^{-j}t-k)dt \quad (4)$$

III. Experimental Results

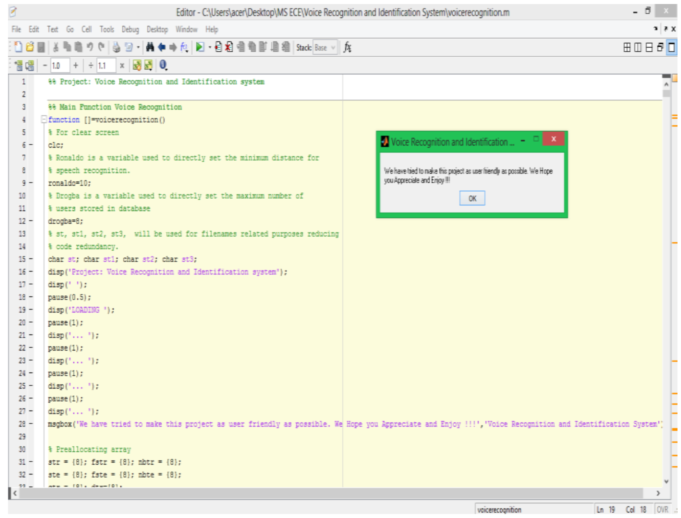

Fig. 3. M-file programming environment in Matlab.

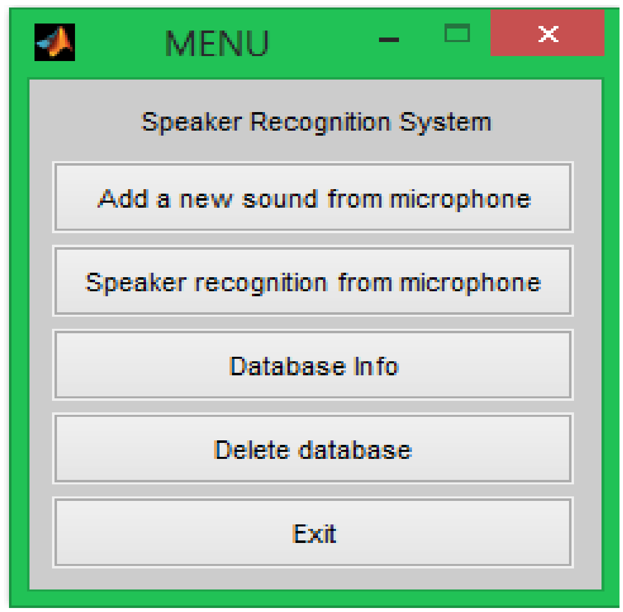

Fig. 4. Developed GUI human-machine interface.

Running the m-file of the Matlab program, the system then runs the graphical user interface (GUI) showing the options for the speaker recognition system. The GUI also pops-out the user with a welcome message. In Fig. 4 then illustrates the GUI interface, which shows the following options: Adding the new sound from microphone, the speaker recognition from microphone, database information, the delete database, and also exiting the system. By selecting the 'Add a new sound from microphone' option, the program shall ask the user to insert a class number or sound ID that shall be used by the program for recognition purposes. This is shown in Fig. 4. The ID shall then be tagged together with the corresponding recorded voice. The program shall also ask the user to input the necessary recording parameters that shall be

ISSN: 2277 – 1581
01 August 2015

used for recording as it is shown in Fig. 5. The sampling frequency and the bits per sample shall also be asked together with the total duration of the recording. The longer the recording time the bigger the file size for the recorded voice.

```
Project: Voice Recognition and Identification system

LOADING
...
...
...
...

> 10: Test with other speech files

> 10: Test with other speech files
Insert a class number (sound ID) that will be used for recognition:10
```

Fig. 5. Matlab program is waiting for voice to be entered for recognition.

```
Project: Voice Recognition and Identification system

LOADING
...
...
...
...

> 10: Test with other speech files

> 10: Test with other speech files
Insert a class number (sound ID) that will be used for recognition:10
The following parameters will be used during recording:
Sampling frequency22050
Bits per sample8
Insert the duration of the recording (in seconds):5
```

Fig. 6. Duration for recording of voice (seconds).

After the recording time is inputted, the GUI will then pop-out to indicate that the sound is added in the database. This is shown in the Fig. 7. Thus, after a voice recorded is added and recorded by the speaker, the user can then now experimentally test the system whether the program can recognize him\her together with its tagged corresponding Sound ID. It shall ask also the same parameters like the sampling frequency, bits per sample, and the time duration. As shown in Fig. 9. the system shall prompt the user to enter another voice after the recording time has been set, which in this case is 8 seconds (Fig. 10). During these 8 seconds, the user shall have to record his/her voice until such time that the recording time is fully stopped as it is shown in Fig. 11. After the recording, the system shall compute the discrete wavelet transform (DWT) coefficients of the voice recorded and compare the computed values to the DWT coefficients of existing voice database. A GUI pops-out which in this case indicates that the voice is recognized successfully together with its corresponding Sound ID. This is shown in Fig. 11. The system also shows the linear and logarithmic power spectrum graph of the sampled voices. As it is shown in Fig. 12., the illustrative graph visualizes the power spectrum and logarithmic power spectrum

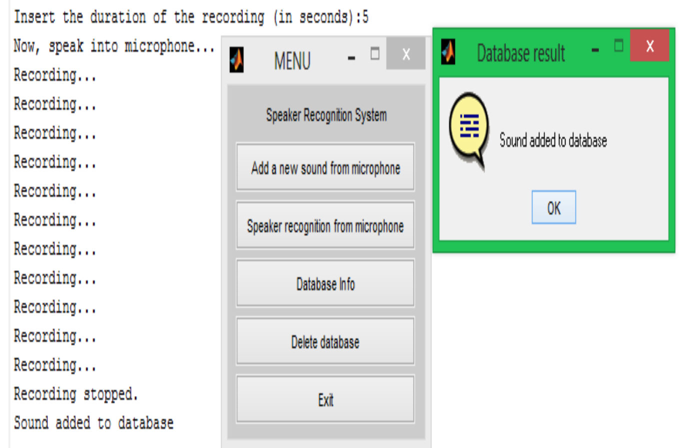

Fig. 7. Recorded voice successfully added to the database.

```
> 10: Test with other speech files
Insert a class number (sound ID) that will be used for recognition:01
The following parameters will be used during recording:
Sampling frequency22050
Bits per sample8
Insert the duration of the recording (in seconds):
```

Fig. 8. Entering the time duration for voice recording (seconds).

```
Sound added to database
fx Insert the duration of the recording (in seconds):8
Start Waiting for input
```

Fig. 9. Entering another voice for wavelet computation.

```
Insert the duration of the recording (in seconds):8
Now, speak into microphone...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording...
Recording stopped.
```

Fig. 10. Actual recording of voice while user is speaking in microphone

ISSN: 2277 – 1581
01 August 2015

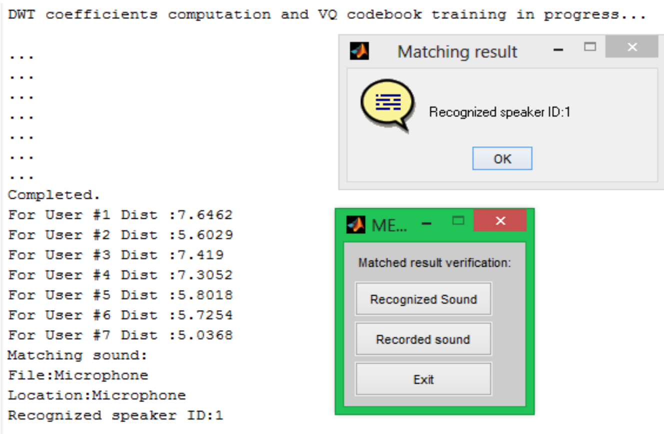

Fig. 11. System successfully recognized the voice using DWT.

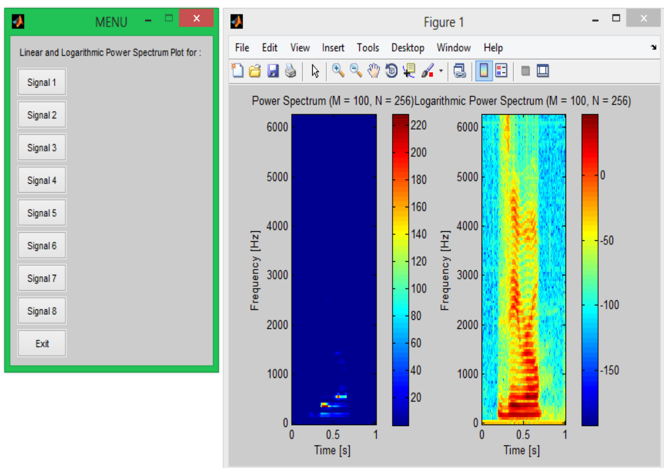

Fig. 12. Power spectrum and logarithmic power spectrum.

IV. Conclusion

In this paper, the speaker voice recognition using discrete wavelet transform (DWT) is then presented. Discrete wavelet transform is used for speaker voice recognition and is able to distinguish the different properties of the voices whether high frequency low amplitude spectral components or whether low frequency large amplitude spectral components.

Experimental studies have been carried out to verify the effectiveness of the proposed scheme. The simulate results have confirmed that the system recognizes the human speaker voice effectively. A wavelet-based voice recognition system in this paper can therefore indeed be successfully implemented.

Acknowledgements

The researchers would like to thank Adamson University for the support and financial grant provided for publication of this project.

References

i. *B.T. Tan, M. Fu, A. Spray, and P. Dermody. "The use of wavelet transform for phoneme recognition," in Proceedings of the 4th International Conference of Spoken Language Processing. vol. 4, Philadelphia, USA, pp. 2431 – 2434. 1996.*

ii. *M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi. Matlab Wavelet Tool Box. The MathWorks Inc., 2000. pp. 795.*

iii. *G. Tzanetakis, G. Essl, and P. Cook. "Audio analysis using the discrete wavelet transform". Organized Sound, vol. 4., no. 3., 2000.*

iv. *S. Tamura, and A. Waibel. "Noise reduction using connectionist models" in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1., pp. 553 – 556. 1988.*

v. *D. Ning. "Developing an isolated word recognition system in Matlab" MATLAB Digest, The MathWorks Inc. 2009.*

vi. *J. Tebelskis. "Speech recognition using neural networks". School of Computer Science, Carnegie Mellon University. 1995.*

vii. *J. C. Principe, and R. C. Dorf. "Artificial neural networks" (extract), The Electrical Engineering Handbook. Ed. Boca Raton: CRC Press LLC. 2000*

viii. *[Online]. Available: http://www.learnartificialneuralnetworks.com/speechrecognition.html*

ix. *[Online]. Available: http://www.imagesco.com/articles/hm2007/SpeechRecognitionTutorial02.html*

x. *C.T. Hsieh, E. Lai, and Y.-C. Wang. "Robust Speaker Identification System Based on Wavelet Transform and Gaussian Mixture Model". Journal of Information Science and Engineering, Vol. 19, pp. 267 - 282, 2003.*